

Um Modelo para Otimizar a Rede de Coleta do IBGE

Eduardo Leoni

eduardo.leoni@ibge.gov.br

IBGE - Superintendência Estadual da Bahia

Introdução

- O Brasil apresenta território vasto e unidades de coleta são distribuídas de forma heterogênea
- O IBGE enfrenta desafios logísticos e altos custos na coleta de dados
- Há crescente dificuldade para compor orçamento das pesquisas
- **Proposta de hoje: redução de custos via realocação de Unidades de Coleta (UCs, e.g. UPAs) às agências**

O Problema de Alocação de UCs a Agências

**Precisamos decidir,
simultaneamente:**



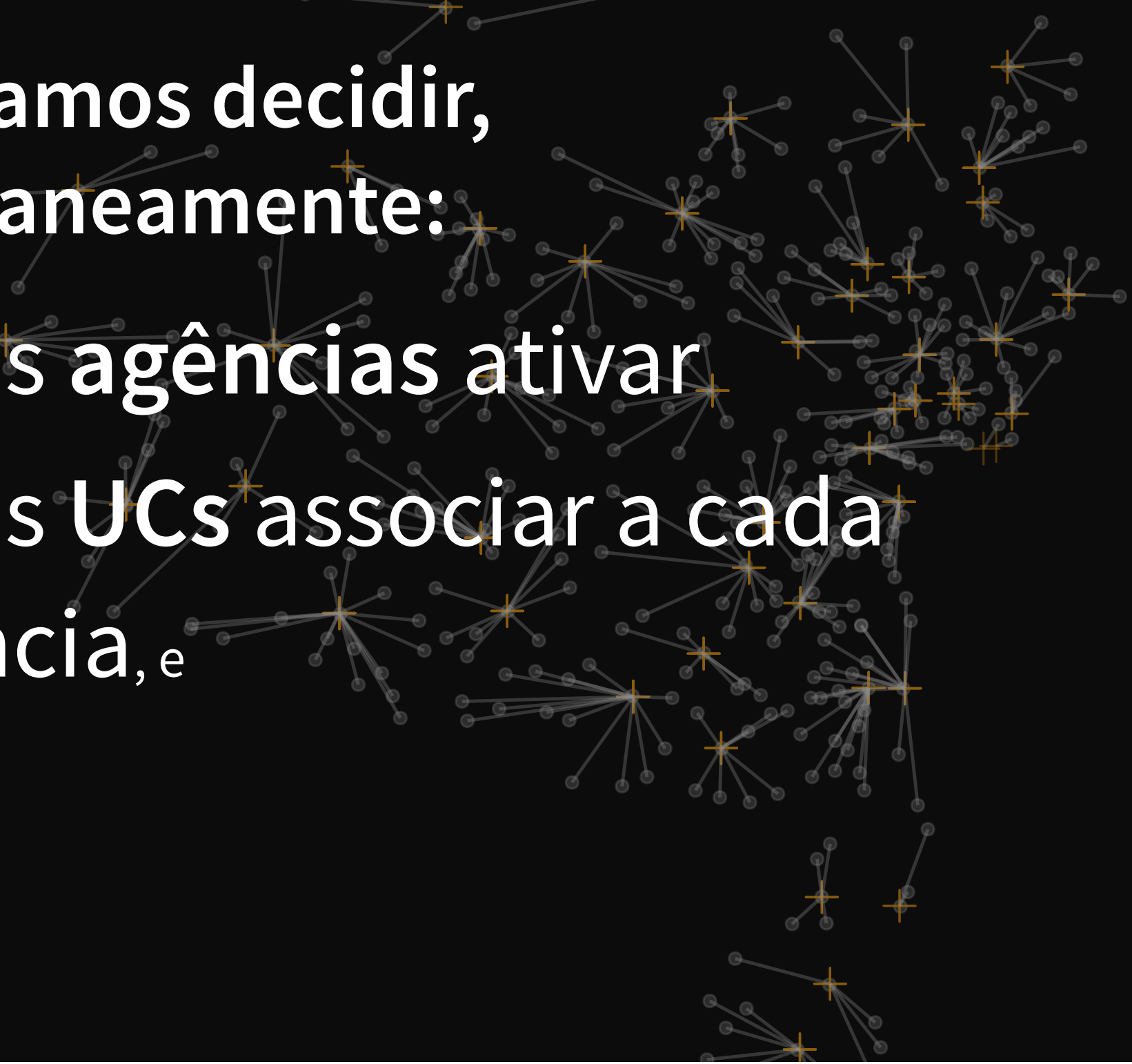
Precisamos decidir,
simultaneamente:

1. Quais agências ativar



Precisamos decidir,
simultaneamente:

1. Quais agências ativar
2. Quais UCs associar a cada agência, e



Precisamos decidir,
simultaneamente:

1. Quais agências ativar
2. Quais UCs associar a cada agência, e
3. Quantos entrevistadores contratar por agência

De forma a minimizar os custos de coleta.



Uma nova ferramenta: **orce**

- **O**timização de **R**edes de **C**oleta **E**statística
- Desenvolvido em R para otimizar a alocação de:
 - Unidades Primárias de Amostragem (UPAs)
 - Escolas (para pesquisas educacionais)
 - Municípios
- Utiliza programação linear inteira mista
- Objetivo: Minimizar os custos totais de coleta, tendo como dada a rede atual de agências do IBGE.

Como encontrar a melhor alocação?

- **Problema Clássico de Referência:** Minimizar custos logísticos na distribuição entre armazéns e consumidores.
Baseado em “**The Warehouse Location Problem**”, por Dirk Schumacher
- **Estrutura:**
 - Armazéns (Agências)
 - Consumidores (UCs)
 - Custos de transporte entre **todos** os Armazéns e **todos** os consumidores (entre Agências e UCs)
 - Número de entrevistadores por agência (**novidade!**)

Formulação matemática

Começamos com um conjunto de UCs $U = \{1 \dots n\}$ e um conjunto de agências $A = \{1 \dots m\}$ que poderiam ser ativadas.

Utilizaremos três variáveis de decisão:

- $x_{i,j}$: uma variável binária que é 1 se a UC i for alocada à agência j , e 0 caso contrário.
- y_j : uma variável binária que é 1 se a agência j for selecionada para realizar a coleta, e 0 caso contrário.
- w_j : uma variável inteira com o número de entrevistadores a contratar na agência j .

minimizar: $\sum_{i=1}^n \sum_{j=1}^m \text{custo_deslocamento}_{i,j} \cdot x_{i,j} +$

$$\sum_{j=1}^m \text{custo_fixo}_j \cdot y_j +$$

$$\sum_{j=1}^m \text{custo_por_entrevistador}_j \cdot w_j$$

sujeito a:

$$\sum_{j=1}^m x_{i,j} = 1 \quad i = 1, \dots, n$$

$$\sum_{i=1}^n x_{i,j} \cdot \frac{\text{dias_coleta}_{i,j}}{\text{dias_max}} \leq w_j \quad j = 1, \dots, m$$

$$x_{i,j} \leq y_j, \quad i = 1, \dots, n \quad j = 1, \dots, m$$

$$x_{i,j} \in \{0, 1\} \quad i = 1, \dots, n \quad j = 1, \dots, m$$

$$y_j \in \{0, 1\} \quad j = 1, \dots, m$$

$$w_j \in \{0, 1, 2, \dots\} \quad j = 1, \dots, m$$

Recursos Principais do **orce**

- Otimização da alocação de UCs a agências
- Otimização do número de entrevistadores por agência
- Auxílio no cálculo de distâncias entre UCs e agências
- Relatórios com cálculo detalhado da previsão dos custos de coleta por UC e por Agência
- Auxílio no cálculo dos custos de treinamento.

Colocando em prática: Fatores considerados na Otimização

Custos de deslocamento

Distâncias/tempo de deslocamento a calcular: Número de Unidades de Coleta \times Número de Agências.

Distâncias/tempo de deslocamento disponíveis para (quase) todos os municípios e agências do IBGE no pacote. Foram calculados com o OpenStreetMap.

Diárias de campo

Cálculo de diárias

Devidas quando a unidade de coleta está fora:

- a. da jurisdição da agência; e
- b. da microrregião da agência; e
- c. da região metropolitana da agência; ou
- d. quando há necessidade de pernoite.

A necessidade de pernoite pode ser estimada utilizando o tempo de deslocamento entre a unidade de coleta e cada agência.

Custos com entrevistadores

- Remuneração
- Custos com treinamento: deslocamento da agência de origem até a agência de treinamento + diárias (se houver)

Restrições e características específicas das agências

- Número máximo de entrevistadores,
- Custos fixos diferenciados.

Função para alocação de UCs: **orce**

Demo

- Unidade de Coleta: Municípios

```
1 head(ucs_municipios)%>%gt::gt()
```

uc	municipio_codigo	agencia_codigo	dias_coleta	viagens	data
2901957	2901957	291760700	10	1	1
2904308	2904308	291760700	10	1	1
2909505	2909505	291760700	10	1	1
2914208	2914208	291760700	10	1	1
2916708	2916708	291760700	10	1	1
2916906	2916906	291760700	10	1	1

```
1 head(agencias)%>%gt::gt()
```

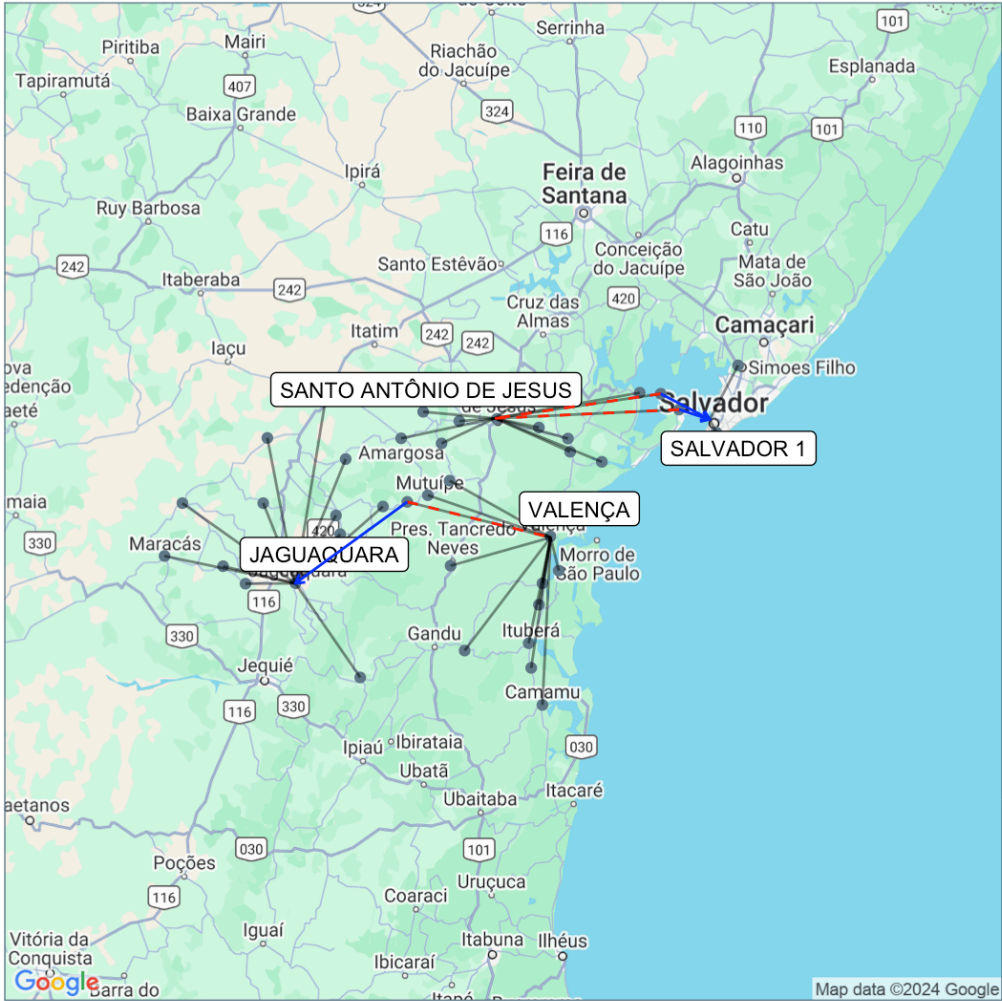
agencia_codigo	custo_fixo	n_entrevistadores_agencia_max
291760700	0	Inf
292740800	0	Inf
292870300	0	Inf
293290300	0	Inf

```
1 head(dists) %>%gt::gt()
```

agencia_codigo	municipio_codigo	distancia_km	duracao_horas	diaria_municipio	diaria_pernoite	uc
290070200	2901007	230,90	3,08	TRUE	TRUE	2901007
290320100	2901007	693,83	9,70	TRUE	TRUE	2901007
290390400	2901007	576,08	8,99	TRUE	TRUE	2901007
290460500	2901007	336,48	5,20	TRUE	TRUE	2901007
290490200	2901007	113,82	1,89	TRUE	TRUE	2901007
290570100	2901007	214,62	3,43	TRUE	TRUE	2901007

```
1 res <- orce(ucs = ucs_municipios,  
2           agencias = agencias,  
3           dias_coleta_entrevistador_max = 20,  
4           distancias_ucs = dists)  
5 str(res)  
6 #> List of 5  
7 #> $ resultado_ucs_otimo      : tibble [41 x  
8 #>   ..$ uc                   : chr [1:41] "2  
9 #>   ..$ agencia_codigo       : chr [1:41] "2  
10 #>   ..$ agencia_codigo_jurisdicao: chr [1:41] "2  
11 #>   ..$ distancia_km        : num [1:41] 57  
12 #>   ..$ duracao_horas        : num [1:41] 1,
```


	Unidades de Coleta	Custo Total Otimo	Custo Total Jurisdicao	Total Diarias Otimo	Total Diarias Jurisdicao	Custo Combustivel Otimo	Custo Combustivel Jurisdicao	Entrevistadores Otimo	Entrevistadores Jurisdicao
Jaguaquara	15 (14 + 1)	R\$11.662	R\$10.584	0	0	R\$9.142	R\$8.286	8	7
Salvador 1	4 (2 + 2)	R\$1.531	R\$545	0	0	R\$1.011	R\$433	2	1
Santo Antônio de Jesus	11 (13 + 0 -2)	R\$6.036	R\$10.765	0	10	R\$4.708	R\$5.929	6	7
Valença	11 (12 + 0 -1)	R\$10.462	R\$11.663	10	10	R\$5.701	R\$6.646	6	6
Total Superintendência	—	R\$29.692	R\$33.558	10	19	R\$20.563	R\$21.295	22	21



municipio_nome	agencia_nome	agencia_nome_jurisdicao	custo_diarias	custo_diarias_jurisdicao	custo_combustivel	custo_combustivel_juri
Itaparica	SALVADOR 1	SANTO ANTÔNIO DE JESUS	0	3182,5	357,72	11
Jiquiriçá	JAGUAQUARA	VALENÇA	0	0,0	855,84	94
Vera Cruz	SALVADOR 1	SANTO ANTÔNIO DE JESUS	0	0,0	220,20	110

Alocação por UC(UPA): Ideal para pesquisas com equipes exclusivas ou com limitação no número de equipamentos (e.g. PeNSE, POF).

Alocação por municípios (ou outro nível):

alocar_por

- Realiza a alocação otimizada de municípios (ou outro variável para agregação) às agências.
- Ou seja, as UCs de um mesmo município são alocadas integralmente para uma mesma agência. **O custo de deslocamento agência-município é a soma dos custo de deslocamento entre cada UCs de um município e a agência candidata.**
- Permite maior “sinergia” da coleta entre as várias pesquisas em campo.

Alocação por período de coleta: **data**

- **data** é uma coluna nos dados sobre cada UC (e.g. UPA)
- Restrição: a carga de trabalho deve ser compatível com o número de entrevistadores para **todos** os períodos de coleta

$$\sum_{i=1}^n x_{i,j,t} \cdot \frac{\text{dias_coleta}_{i,j,t}}{\text{dias_max}} \leq w_j$$

$$j = 1, \dots, m$$

$$t = 1, \dots, p$$

**Poderíamos ter feito
melhor? A coleta da
PNDS na SES-BA**

Especificidades da PNDS

- Equipes mistas (homem/mulher), então mínimo de 2 entrevistadores por equipe
- Custos altos de treinamento presencial
- Há um único período de coleta, de quatro meses

“Otimização” da Rede de Coleta

- Análise feita **manualmente**, após calcular as distâncias entre as UPAs e as agências mais próximas.
- Nem todas as sugestões foram implementadas. Mas foram reduzidos o número de agências e o número de treinandos.
- Nos próximos slides veremos se o **orç**ce teria encontrado alocações mais eficientes para a coleta.

Parâmetros para o cálculo utilizando o **orce** por município:

- 60 dias de trabalho por entrevistador na pesquisa (4 meses)
- 8 dias de coleta por UPA (são dois entrevistadores, esse valor é o total)
- 2 viagens, no caso de diárias inteiras, e 8 viagens nos casos sem diárias (ou com meias-diárias).
- São sugeridas alterações de jurisdição somente quando a economia proporcionada for maior que R\$ 100.

Parâmetros (2):

- **Todas as viagens são feitas por automóvel**, o custo por litro de combustível é de **R\$ 6**, o automóvel percorre **10 quilômetros** com um litro. Cada hora de viagem tem custo adicional equivalente a **R\$ 0**.
- Viagens (ida) com mais de **1,5** horas de duração fazem jus a diárias.
- A diária tem o valor de **R\$ 335**, e é paga nos casos de pernoite ou quando a UPA é de fora da **jurisdição/microrregião/região metropolitana** da agência.
- Remuneração por entrevistador por mês: **R\$ 1300** .
- Número de meses de coleta: **4**
- Mínimo de **2** entrevistadores por agência. (Eram necessários pelo menos um entrevistador homem e uma entrevistadora

Parâmetros (3):

- 5 dias de treinamento por entrevistador
- Agência(s) de treinamento: **Feira de Santana e Salvador**
- diarias_entrevistador_max = **Inf**
- agencias_treinadas = **NULL**
- distancias_agencias: **distancias_agencias_osrm** dados com as distâncias rodoviárias entre todas as agências de cada UF, disponíveis no pacote

Parâmetro	Valor
horas_viagem_pernoite	1,5
custo_hora_viagem	0
custo_litro_combustivel	6
kml	10
valor_diaria	335
adicional_troca_jurisdicao	100
remuneracao_entrevistador_mes	1300
n_entrevistadores_min	2
viagens	2
dias_coleta	8
dias_treinamento	5
dias_coleta_entrevistador_max	60
solver	cbc
max_time	30
rel_tol	0,005
n_meses_coleta	4

Resultado da Otimização

Tempo máximo de execução: 20 segundos.

	Jurisdição	Ótimo Ex Rem	Ótimo c Rem	Projeção	Realizado
N Agencias	-	-	-	-	-
Entrevistadores	-	-	-	-	-
Custo Treinamento	-	-	-	-	-
Custo Combustivel	-	-	-	-	-
Custo Diarias	-	-	-	-	-
Custo Deslocamento	-	-	-	-	-
Custo Excluindo Remuneracao	-	-	-	-	-
Custo Incluindo Remuneracao	-	-	-	-	-
Obs: Estamos incluindo no custo de deslocamento aqui somente a combustível e diárias, ou seja, não há custo por hora de viagem.					

	Jurisdição	Ótimo Ex Rem	Ótimo c Rem	Projeção	Realizado
N Agencias	47	-	-	-	-
Entrevistadores	96	-	-	-	-
Custo Treinamento	194,437	-	-	-	-
Custo Combustivel	98,173	-	-	-	-
Custo Diarias	115,575	-	-	-	-
Custo Deslocamento	213,748	-	-	-	-
Custo Excluindo Remuneracao	408,185	-	-	-	-
Custo Incluindo Remuneracao	907,385	-	-	-	-
Obs: Estamos incluindo no custo de deslocamento aqui somente a combustível e diárias, ou seja, não há custo por hora de viagem.					

	Jurisdição	Ótimo Ex Rem	Ótimo c Rem	Projeção	Realizado
N Agencias	47	33	-	-	-
Entrevistadores	96	69	-	-	-
Custo Treinamento	194,437	135,236	-	-	-
Custo Combustivel	98,173	108,648	-	-	-
Custo Diarias	115,575	115,240	-	-	-
Custo Deslocamento	213,748	223,888	-	-	-
Custo Excluindo Remuneracao	408,185	359,124	-	-	-
Custo Incluindo Remuneracao	907,385	717,924	-	-	-
Obs: Estamos incluindo no custo de deslocamento aqui somente a combustível e diárias, ou seja, não há custo por hora de viagem.					

	Jurisdição	Ótimo Ex Rem	Ótimo c Rem	Projeção	Realizado
N Agencias	47	33	22	-	-
Entrevistadores	96	69	46	-	-
Custo Treinamento	194,437	135,236	87,356	-	-
Custo Combustivel	98,173	108,648	107,566	-	-
Custo Diarias	115,575	115,240	212,892	-	-
Custo Deslocamento	213,748	223,888	320,458	-	-
Custo Excluindo Remuneracao	408,185	359,124	407,814	-	-
Custo Incluindo Remuneracao	907,385	717,924	647,014	-	-
Obs: Estamos incluindo no custo de deslocamento aqui somente a combustível e diárias, ou seja, não há custo por hora de viagem.					

	Jurisdição	Ótimo Ex Rem	Ótimo c Rem	Projeção	Realizado
N Agencias	47	33	22	41	-
Entrevistadores	96	69	46	75	-
Custo Treinamento	194,437	135,236	87,356	153,279	-
Custo Combustivel	98,173	108,648	107,566	103,988	-
Custo Diarias	115,575	115,240	212,892	128,472	-
Custo Deslocamento	213,748	223,888	320,458	232,460	-
Custo Excluindo Remuneracao	408,185	359,124	407,814	385,739	-
Custo Incluindo Remuneracao	907,385	717,924	647,014	775,739	-
Obs: Estamos incluindo no custo de deslocamento aqui somente a combustível e diárias, ou seja, não há custo por hora de viagem.					

	Jurisdição	Ótimo Ex Rem	Ótimo c Rem	Projeção	Realizado
N Agencias	47	33	22	41	-
Entrevistadores	96	69	46	75	-
Custo Treinamento	194,437	135,236	87,356	153,279	-
Custo Combustivel	98,173	108,648	107,566	103,988	-
Custo Diarias	115,575	115,240	212,892	128,472	-
Custo Deslocamento	213,748	223,888	320,458	232,460	-
Custo Excluindo Remuneracao	408,185	359,124	407,814	385,739	-
Custo Incluindo Remuneracao	907,385	717,924	647,014	775,739	-
Custo Incluindo Remuneracao Rel	0	-21	-29	-15	-
Obs: Estamos incluindo no custo de deslocamento aqui somente a combustível e diárias, ou seja, não há custo por hora de viagem.					

	Jurisdição	Ótimo Ex Rem	Ótimo c Rem	Projeção	Realizado
N Agencias	47	33	22	41	41
Entrevistadores	96	69	46	75	77
Custo Treinamento	194,437	135,236	87,356	153,279	149,717
Custo Combustivel	98,173	108,648	107,566	103,988	79,928
Custo Diarias	115,575	115,240	212,892	128,472	148,424
Custo Deslocamento	213,748	223,888	320,458	232,460	228,352
Custo Excluindo Remuneracao	408,185	359,124	407,814	385,739	378,069
Custo Incluindo Remuneracao	907,385	717,924	647,014	775,739	778,469
Custo Incluindo Remuneracao Rel	0	-21	-29	-15	-14
Obs: Estamos incluindo no custo de deslocamento aqui somente a combustível e diárias, ou seja, não há custo por hora de viagem.					

	Unidades de Coleta	Custo Total Otimizado	Custo Total Jurisdiccional	Total Diárias Otimizado	Total Diárias Jurisdiccional	Custo Combustível Otimizado	Custo Combustível Jurisdiccional	Entre		
Alagoinhas	15 (11 + 4)	R\$29.219	R\$19.678	30	8	R\$5.243	R\$3.239			
Barreiras	14 (9 + 5)	R\$40.272	R\$26.420	56	22	R\$5.787	R\$3.324			
Camaçari	14 (11 + 3)	R\$17.356	R\$16.475	0	0	R\$3.492	R\$2.611			
Conceição do Coité	8 (7 + 1)	R\$16.655	R\$16.110	0	0	R\$2.657	R\$2.112			
Eunápolis	10 (3 + 7)	R\$23.428	R\$15.181	15	0	R\$3.309	R\$87			
Feira de Santana	30 (20 + 10)	R\$37.800	R\$18.422	30	0	R\$6.950	R\$2.822			
Guanambi	15 (11 + 4)	R\$38.015	R\$26.495	52	22	R\$5.320	R\$3.850			
Irecê	15 (5 + 10)	R\$45.446	R\$16.208	75	0	R\$5.687	R\$1.574			
Itabuna	13 (6 + 7)	R\$32.084	R\$21.149	38	15	R\$4.935	R\$1.538			
Jacobina	13 (6 + 7)	R\$34.981	R\$19.404	45	8	R\$5.616	R\$2.602			
1–10 of 23 rows						Previous	1	2	3	Next

Agências excluídas: Bom Jesus da Lapa, Brumado, Cachoeira, Cipó, Cruz das Almas, Esplanada, Euclides da Cunha, Ibotirama, Ilhéus, Ipiaú, Ipirá, Itaberaba, Itamaraju, Itapetinga, Jequié, Jeremoabo, Livramento de nossa Senhora, Morro do Chapéu, Porto Seguro, Remanso, Santa Rita de Cássia, São Francisco do Conde, Seabra, Serrinha, Xique Xique

Conclusão: Benefícios da Otimização:

- **Melhoria no Planejamento:**
 - Calcula o custo da pesquisa (diárias, treinamento, combustível, e remuneração).
 - Facilita o planejamento e gestão da coleta de dados
- **Otimização de Recursos:**
 - Otimiza a alocação de entrevistadores e UCs entre as agências.
 - Permite maior equilíbrio na carga de trabalho das agências

Outras funcionalidades

- **Alocação de UCs/municípios levando em conta o cronograma das pesquisas** Na POF, por exemplo, podemos considerar como “Unidade de Coleta” as UPAs por período teórico.
- **Pontos representativos das UPAs** Em vez da centroides dos setores ou UPAs, utiliza o CNEFE e seleciona o domicílio na região de maior densidade (em número de domicílios).
- Função `calcula_distancias` utiliza dados do OpenStreetMaps com a Open Source Routing Machine (OSRM) para calcular distâncias de uma matriz de pontos.

(Possíveis) Próximos Passos

- Alocação intramunicipal com dados de transporte público
- Utilizar algoritmo de “traveling salesman problem” para alocação ótima de grupos de UCs
- Otimizar planos de alteração da jurisdição das agências.

Obrigado!

Pacote disponível no [Github](#)

Comentários bem vindos! eduardo.leoni@ibge.gov.br

Detalhes Técnicos

Restrições opcionais

- Restrições de capacidade por período de coleta (e.g. mês)
- Número máximo de dias de coleta de cada agência
- Número mínimo e máximo de entrevistadores por agência ativa
- Máximo de diárias por entrevistador

Implementação do modelo utilizando o pacote **ompr**

```
1 model <- ompr::MIPModel() |>
2   # 1 sse uc i vai para a agencia j
3   ompr::add_variable(x[i, j], i = 1:n, j = 1:m,
4   # 1 sse agencia j ativada
5   ompr::add_variable(y[j], j = 1:m, type = "bin
6   # trabalhadores na agencia j
7   ompr::add_variable(w[j], j = 1:m, type = n_en
8   # minimizar custos
9   ompr::set_objective(
10     ompr::sum_over(transport_cost_i_j[i, j] * x
11     ompr::sum_over((agencias_t$custo_fixo[j])
12     w[j] * ({remuneracao_entrevist
```

Pontos representativos das UPAs

- Não utilizamos a centroide dos setores ou UPAs
- Utilizamos, em vez disso, o CNEFE (público).
- Para cada UPA, utilizando a função `ponto_densidade`, calculamos a densidade de domicílios para cada domicílio da UPA. O domicílios com a maior densidade calculada é considerado o ponto representativo do domicílio.
- Esse procedimento diminui a possibilidade de calcular distâncias de agências a pontos vazios ou inacessíveis nos setores/UPAs.

Escolhendo qual **so_lver** utilizar

- Sugerimos utilizar o **c_bc** CBC (COIN-OR branch and cut). Mas é um pouco difícil de instalar.
- O **h_ighs** às vezes é mais rápido, mas falha nos problemas mais difíceis.
- O **g_lp_k** é o mais fácil de instalar, mas é também o mais lento.